Optimization of the $X(3872) \rightarrow J/\psi \pi^+ \pi^-$ selection at LHCb

Annabelle Chuinard
annabelle.chuinard@epfl.ch

October 13, 2011

Contents

1 Background and Signal .................................................. 3
  1.1 Events Samples .................................................. 3
  1.2 Normalization .................................................. 4

2 Preselection tuning ....................................................... 5
  2.1 Preselection .................................................... 6
  2.2 Events distribution for each variable ......................... 7

3 Optimized selection ...................................................... 8
  3.1 Manual independent cuts ....................................... 8
      3.1.1 Settings .................................................. 8
      3.1.2 Results .................................................. 8
      3.1.3 Application .............................................. 9
  3.2 Fisher discriminant method for optimization .................. 11
      3.2.1 Theory .................................................. 11
      3.2.2 Settings ................................................ 12
      3.2.3 Application ............................................ 17
  3.3 Rectangular cuts with TMVA .................................... 19
      3.3.1 Operational principles .................................. 19
      3.3.2 Results .................................................. 19
      3.3.3 Application ............................................ 20
The \( X(3872) \) meson has been discovered by the Belle Experiment [1] in 2003 and quickly confirmed by CDF [2] in 2004 and BaBar collaboration [3] in 2007. Although the \( X(3872) \) properties has been well studied: for instance, its mass has been measured by CDF and LHCb [4] and its quantum number restrained to \( 1^{++} \) and \( 2^{-+} \) [5], its nature remains enigmatic. Different models have been proposed, all with a limited success.

It is possible for \( X(3872) \) to be a simple charmonium state i.e. a \( c\bar{c} \) meson. The candidate we have for this is the \( \eta_{c2}(1D) \) meson [6]. However, the mass for this state is predicted to be different from the observed \( X(3872) \) mass. Besides, it has been shown by CDF that the mass distribution of the two pions produced was dominated by the \( \rho^0 \) resonance, a channel which would be isospin violating for charmonium and thereby, would discredit this model.

Another singular possibility is to consider the idea of Maiani [7] that the \( X(3872) \) would be a tetraquark state. This model predicts the existence of two distinct neutral states with a difference in mass of \( \Delta m \approx 8 \pm 3 \text{MeV}/c^2 \). This hasn’t been corroborated by the experiment (CDF [8], Belle [9]). Indeed, CDF sets a limit of the mass difference to 3.6 MeV/c\(^2\), disfavouring the tetraquark model.

The most popular model to date is a deuteron-like molecule. It can actually explain all the properties and decay modes of the \( X(3872) \) mesons. Whether the proposition of bound state is valid crucially depends on the question whether the \( X(3872) \) is heavier or lighter than the sum of the masses it is made up of. In other words, we should verify that the mass is below the \( D^{*0}\bar{D}^0 \) threshold. For that purpose, a precise determination of this threshold and of the \( X(3872) \) mass are needed.

This report presents different optimization methods of the selection of the \( X(3872) \) state using the \( X(3872) \rightarrow J/\psi\pi^+\pi^- \) channel with \( J/\psi \rightarrow \mu^+\mu^- \) at LHCb [10]. In order to select the events really corresponding to this occurrence, we have to optimize the cuts on the observable variables. In other words, we need to choose likely ranges for the values of the channel-describing variables. A selection will be made among, for instance, the total momentum of the daughter particles, their transverse momentum, the precision of the track they might leave in the detector or the \( Q^1 \) of the decay in order to maximize the significance of our signal. These selections are built relying on simulated signal and same-sign background. This optimized selection will be applied on the data.

\[ 1Q = M_{\mu\mu\pi\pi} - M_{\mu\mu} - M_{\pi\pi} \] where \( M_{\mu\mu\pi\pi}, M_{\mu\mu} \) and \( M_{\pi\pi} \) are the reconstructed masses before any mass constraint.
We will aim at finding the most efficient selection on the variables in order to allow the best separation between noise-related and signal-related events. This requirement is satisfied when the significance is maximized. In this case, the significance is defined by:

\[ \mathcal{S}_{\text{real}} = \frac{S}{\sqrt{S + B}} \]

where:
- \( S \) | the number of signal events
- \( B \) | the number of background events

1 Background and Signal

1.1 Events Samples

To do the analysis of the \( X(3872) \) mass, we use two data samples. The first sample consists in Monte-Carlo generated \( X(3872) \rightarrow J/\psi \pi^+ \pi^- \) events. The second sample called ”same-sign” contains candidates from data which satisfy the same requirements used for \( X(3872) \) selection but where the selected pions have the same electric charge.

- **Signal sample:**

  The production of \( X(3872) \) is recreated thanks to a Monte Carlo sample at an available energy in the center-of-mass of \( \sqrt{s} = 7 \) TeV [10]. We assumed that the two pions are coming from a \( \rho^0 \) meson decay. After preselection, there is a total of 74794 events of this type. The two pions and the two muons produced are required to be within the LHCb acceptance\(^2\). Moreover, the \( X(3872) \) particles are produced with \( J^{PC} = 1^{++} \) quantum numbers.

- **Background sample:**

  The background is described with 117149 selected events which contain each one \( J/\psi \) and two pions of the same sign. The preselection procedure is described in section 2. These were recorded with the LHCb detector between July and November 2010. The available energy in the center-of-mass frame is \( \sqrt{s} = 7 \) TeV. The amount of measured data corresponds to an integrated luminosity of 34.13 ± 3.4 pb\(^{-1}\). The sample is separated in two sub-sets of approximately 17 pb\(^{-1}\) recorded with the two possible magnets polarity (upward and downward).

Since the most important contribution to the combinatorial background corresponds to events coming from a bad recombination of \( J/\psi \) with two same-charged pions (\( J/\psi\pi^+ \pi^+ \) and \( J/\psi\pi^- \pi^- \)), we can use the same-sign events as a description of the combinatorial background.

\(^2\)This condition implies that the all four daughter particles are travelling in a [10,400] mrad angular range before the magnet.
Indeed, not all of the background is coming from a misassociation of two pions with a \( J/\psi \). Misreconstructed \( J/\psi \rightarrow \mu\mu \) are contributing to the total effective background as well. This is the reasons why we can observe a small shift between effective backgrounds and the same-sign one (Fig 1).

### 1.2 Normalization

Unlike the signal-over-background ratio, the significance is sensitive to the yield of signal and noise. Therefore, it is important to take into account the proportion of these two contributions to the final signal.

After a tight selection describe in [4], the number of events involving a production of \( X(3872) \) in the considered decay is 585 while there are 14736 \( X(3872) \)-events generated with Monte-Carlo simulation. In order to work at the same luminosity, we need then to divide our simulated events by a factor of 24.63.

Since the same-sign events underestimate the combinatorial background, we compensate this effect by multiplying the distribution of the SS-events by a factor of 1.17 so as to faithfully reproduce the background behavior (since the shape of the same-sign and opposite signs distributions are the same).

When these adjustments to the real data are completed, we have 585 events of signal and 5013 events of background. With this well chosen normalisation, the significance defined at eq. 1, is equal to the significance we obtained with simulated and same-signed data samples:

\[
\mathcal{I} = \frac{S_{MC}}{\sqrt{S_{MC} + B_{SS}}}
\]  

(2)

where:

| \( S_{MC} \) | the number of Monte-Carlo simulated signal events reweighted |
| \( B_{SS} \) | the number of same-sign events reweighted |
2 Preselection tuning

The event preselection is performed in three distinct steps. A first preselection is done at the hardware level using trigger algorithms (mainly $J/\psi$ trigger). Then, a selection of all the $J/\psi \rightarrow \mu\mu$ candidates is done. This step is called the stripping. In this study, we used the tight $J/\psi$ line ($J\psi_{2mumu}$ line) from di-muon stream in stripping 12. Finally, $J/\psi\pi\pi$ candidates are built using the DaVinci framework and stored in a ROOT n-tuple.

The variables we will perform cuts on are the following ones:

- $Q$ (Q) defined as followed: $Q = M_{\mu\mu\pi\pi} - M_{\mu\mu} - M_{\pi\pi}$
- $p_{T_J}(\text{PT}_{J\psi})$ the transverse momentum of $J/\psi$
- $p_{T_{\mu_1,2}}$ (pt_{mu1,2}) the transverse momentum of the first (second) muon
- $p_{T_{\pi_1,2}}$ (pt_{pi1,2}) the transverse momentum of the first (second) pion
- $\text{pid}_{k1,2}$ (pidk1,2) the likelihood difference between first (second) pion and electron hypotheses.

In order to reject a large fraction of the background while keeping the signal, a cut is chosen as a ±10 MeV window around the true Monte Carlo mass $M$.

![Figure 1: Invariant mass distribution of $J/\psi\pi^+\pi^-$ (black markers) and same-sign $J/\psi\pi^+\pi^-$ candidates (blue-filled histogram). The fit to the mass distribution is represented with a blue line as detailed in [4]. The huge left peak corresponds to $\psi(2S)$ events and the right one to the $X(3872)$ events. The behavior of same-sign and combinatorial background are similar.](image-url)
2 PRESELECTION TUNING

2.1 Preselection

Table 1 presents the final preselection for $J/\psi \rightarrow \mu \mu$. The muons and the $J/\psi$ produced are required to have a sufficient transverse momentum\(^3\) $p_{TJ/\psi}$, $p_{T\mu_{1,2}}$ (PT_JPSI and pt_mu1,2 in the code) and a good track in the detector.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Requirement</th>
<th>Utility</th>
</tr>
</thead>
<tbody>
<tr>
<td>$J/\psi$ : vertex $\chi^2$/ndof</td>
<td>$&lt; 20$</td>
<td>require $J/\psi$ well reconstructed</td>
</tr>
<tr>
<td>$\mu$ : track fit $\chi^2$/ndof</td>
<td>$&lt; 4$</td>
<td>require both $\mu$ have good tracks</td>
</tr>
<tr>
<td>$p_T$ of $J/\psi$</td>
<td>$&gt; 2$ GeV/c</td>
<td>improve the signal purity</td>
</tr>
<tr>
<td>$p_T$ of $\mu^{\pm}$</td>
<td>$&gt; 0.7$ GeV/c</td>
<td>improve signal purity</td>
</tr>
<tr>
<td>Invariant mass $M_{\mu\mu}$</td>
<td>$\in [3040 ; 3140]$ MeV/c(^2)</td>
<td>reduce $J/\psi$ background</td>
</tr>
</tbody>
</table>

Table 1: Preselection settings for $J/\psi \rightarrow \mu^+\mu^-$

$X(3872)$ candidates are coming from the combination of $J/\psi$ particles with oppositely-charged pions. As summed up in Table 2, these pions are also required to have good tracks, sufficient momentum and transverse momentum and to be discernible from kaons. To constrains the final particles $\pi^+$, $\pi^-$ and $J/\psi$ to come from the same original point, we need to apply a vertex fit on the data. Furthermore, in order to reduce the background versus the signal, it is shrewd to constrains $Q$. The $\mu\mu$ invariant mass is also constrained to the nominal $J/\psi$ mass.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Requirement</th>
<th>Utility</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi$ : track fit $\chi^2$/ndof</td>
<td>$&lt; 15$</td>
<td>require good tracks for the pions</td>
</tr>
<tr>
<td>$J/\psi\pi\pi$ : vertex $\chi^2$/ndof</td>
<td>$&lt; 5$</td>
<td>daughter particles coming from a common vertex</td>
</tr>
<tr>
<td>$p_T$ of $\pi^{\pm}$</td>
<td>$&gt; 0.3$ GeV/c</td>
<td>improve signal purity</td>
</tr>
<tr>
<td>$\Delta \ln \mathcal{L}_{K\pi}$</td>
<td>$&lt; 5$</td>
<td>remove pollutant $K$ distinguishing them from $\pi$</td>
</tr>
</tbody>
</table>

Table 2: Preselection settings for $X(3872) \rightarrow J/\psi(\rightarrow \mu\mu)\pi\pi$

\(^3\)In this report, we will write $p_T$ for the transverse momentum of a particle.
2.2 Events distribution for each variable

Fig. 2 presents the distributions of events as a function of $Q$, $p_{T_{J/\psi}}$, $p_{T_{\pi^+}}$, $p_{T_{\mu^+}}$, $p_{idk_1}$ and $p_{idk_2}$ with the cuts given on Table 3.

![Graphs showing distributions](image)

(a) $Q$ distributions  
(b) $p_{T_{J/\psi}}$ distributions

(c) $p_{T_{\pi^+}}$ distributions  
(d) $p_{T_{\mu^+}}$ distributions

(e) $p_{idk_1}$ distributions  
(f) $p_{idk_2}$ distributions

Figure 2: Events distribution for each variable $Q$, $p_{T_{J/\psi}}$, $p_{T_{\pi^+}}$, $p_{T_{\mu^+}}$, $p_{idk_1}$ and $p_{idk_2}$. In red, the distributions corresponding to same-sign events. In blue, the distribution coming from Monte Carlo simulation. Each distribution is normalized so that the number of events integrated over each variable, between its defined minimum and maximum, is equal to 1.
3 Optimized selection

3.1 Manual independent cuts

3.1.1 Settings

The first idea we applied was to manually determine the appropriate cuts on each variable $Q$, $p_{T_{J/\psi}}$, $p_{T_{v_1}}$, $p_{T_{v_2}}$, pidk$_1$ and pidk$_2$. First, the determination of the optimum cut has been done as a guesstimate for each variable, in order to know approximately where the maximizing value will be located. These cuts are presented on Table 3.

Then a loop over each cut-value, computing the significance for each step, has been implemented. For each loop over the variables, we plotted a curve (significance versus cut value) which reaches its maximum for the optimal value of the cut on the variable. The ROOT -code is presented in section 1 of the appendix. To be consistent, during the loop, the other variables we do not vary on are kept fixed to the values shown in table 3.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Cut</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q$</td>
<td>$&lt; 300$ Mev/c$^2$</td>
</tr>
<tr>
<td>$p_{T_{J/\psi}}$</td>
<td>$&gt; 3500$ Mev/c</td>
</tr>
<tr>
<td>$\min(p_{T_{v_1}}, p_{T_{v_2}})$</td>
<td>$&gt; 500$ MeV/c</td>
</tr>
<tr>
<td>$\min(p_{T_{v_1}}, p_{T_{v_2}})$</td>
<td>$&gt; 1000$ MeV/c</td>
</tr>
<tr>
<td>pidk$_1$</td>
<td>$&lt; 0$</td>
</tr>
<tr>
<td>pidk$_2$</td>
<td>$&lt; 0$</td>
</tr>
</tbody>
</table>

Table 3: First estimation of the optimal cuts

3.1.2 Results

The optimum cuts are presented in Table 4. We can easily see the improvement on $\mathcal{I}$ induced by this simple one-variable optimization. This improvement is calculated in such a way:

$$\text{improvement} = 100 \cdot \frac{\mathcal{I} - \mathcal{I}_{\text{rough}}}{\mathcal{I} + \mathcal{I}_{\text{rough}}}$$

where $\mathcal{I}_{\text{rough}}$ is the significance obtained with the rough cuts chosen in Table 3.

The error on the cuts crucially depends on the histogram binning. Indeed, the more our binning would be thin, the less we have chance to make mistakes on the evaluation of the optimal cut. Depending on the width of the bins we chose, we calculated the maximum error as it appears in Table 4. In figure 3, we present the histograms we plotted to determine the
maximum significance $\mathcal{S}$.

The value corresponding to the maximum significance for $\text{pidk}_1$ and $\text{pidk}_2$ should be the same a priori. The different values obtained are due to the way we compute the value at maximum. As already mentioned, the result is correct at a precision depending on the size of each bin in the distribution histogram. In such a case, it is not impossible to get slightly different values for two variables with the same behavior.

<table>
<thead>
<tr>
<th>Variable</th>
<th>$\mathcal{I}$</th>
<th>$\mathcal{P}$</th>
<th>Optimum cut</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q$</td>
<td>10.0</td>
<td>0.2</td>
<td>175 ± 25</td>
<td>17 %</td>
</tr>
<tr>
<td>$p_{T_{J/\psi}}$</td>
<td>7.87</td>
<td>0.1</td>
<td>3250 ± 50</td>
<td>5.2 %</td>
</tr>
<tr>
<td>$\min(p_{T_{x_1}}, p_{T_{x_2}})$</td>
<td>8.33</td>
<td>0.1</td>
<td>475 ± 25</td>
<td>8.0 %</td>
</tr>
<tr>
<td>$\min(p_{T_{\mu_1}}, p_{T_{\mu_2}})$</td>
<td>7.97</td>
<td>0.1</td>
<td>775 ± 25</td>
<td>5.9 %</td>
</tr>
<tr>
<td>$\text{pidk}_1$</td>
<td>9.53</td>
<td>0.6</td>
<td>-33 ± 1</td>
<td>15 %</td>
</tr>
<tr>
<td>$\text{pidk}_2$</td>
<td>8.09</td>
<td>0.3</td>
<td>-30 ± 2</td>
<td>6.5 %</td>
</tr>
</tbody>
</table>

Table 4: Significance ($S/(\sqrt{S+B})$), purity ($S/B$), optimum cut and improving percentage for each variable

### 3.1.3 Application

We need then to apply some of these cuts on the real data. We notice that all the cuts we have been computing are close to the ones we got in Table 3 except the one on $Q$ (Table 4 shows that the most significant improvement of the significance, 17%, is realized since $Q < 175$ MeV). We will apply only a cut on this variable because it is the one which will give us the most efficient result. Table 5 is giving the results after fitting on the selected data:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Fit value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass</td>
<td>3869.27 ± 1.02 MeV</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>6.43 ± 1.55 MeV</td>
</tr>
<tr>
<td>$N_{\text{events}}$</td>
<td>1248 ± 212</td>
</tr>
<tr>
<td>$\mathcal{P} = S/B$</td>
<td>0.12</td>
</tr>
<tr>
<td>$\mathcal{I}$</td>
<td>11.5</td>
</tr>
</tbody>
</table>

Table 5: Fit results for $X(3872)$ with a manual cut on $Q$

The value of $\mathcal{I}$ is consistent with the improvement computed in Table 4 for a cut on $Q$. The significance drift from the value $\mathcal{I} = 10.5$ to the value $\mathcal{I} = 12$, which corresponds to an increasing of 15%.
Figure 3: Significance as a function of the cut on $Q$, $p_{T_{J/\psi}}$, $\min(p_{T_{\pi_1}}, p_{T_{\pi_2}})$, $\min(p_{T_{\mu_1}}, p_{T_{\mu_2}})$, pidk$_1$ and pidk$_2$
3.2 Fisher discriminant method for optimization

In this section, we will aim at finding the best linear combination of $Q$, $p_{T_{j/p}}$, $p_{T_{u}}$, and $p_{T_{v}}$, allowing us to find the best cut to apply on the data. This will be performed by TMVA (Toolkit for Multivariate Data Analysis with ROOT) [11] using a Fisher’s discriminant method [12].

3.2.1 Theory

The linear discriminant analysis determines an axis in the $n$-dimensional space of the $n$ cutting variables such that, when projecting signal and background outputs upon this axis, they are pushed as far as possible away from each other, while events of each type, signal or background are so far as confined in a close vicinity. The covariance matrix of the discriminating variable space give us an idea of how the variable are linked to each other.

Figure 4 is a graphical 2-dimensional example showing the difference between simple rectangular cuts as implemented at section 3.1 and a cut built from a combination of two variables.

![Diagram of Fisher cuts](image_url)

Figure 4: Comparison for a 2-dimensional space of variables between rectangular cuts and Fisher cuts

Let $x_1, \ldots, x_n$ be the variables which describe our system and let the variables:

- $\bar{x}_i$ be the mean value of the variable $k$ for all entries, $i = 1, \ldots, n$
- $\bar{x}_{S,i}$ be the mean value of the variable $k$ for every entries corresponding to signal
- $\bar{x}_{B,i}$ be the mean value of the variable $k$ for every entries corresponding to signal
- $C$ be the covariance matrix of the sample which can be decomposed in:

$$C = W + B$$
where $W$ is the matrix describing the dispersion of events relative to the means of their own type (background or signal) and $B$ the dispersion of events relative to the means of the whole data. We can then compute the Fisher coefficients, $F_i$:

$$F_i = \frac{\sqrt{S_{MC} B_{SS}}}{S_{MC} + B_{SS}} \sum_{j=1}^{n} W^{-1}_{ij} (\bar{x}_{S,j} - \bar{x}_{B,j})$$

(3)

The Fisher discriminant $F_k$ for event $k$ is given by:

$$F_k = F_0 + \sum_{i=1}^{n} F_i x_i(k)$$

(4)

where $F_0$ centers the sample mean of all $S_{MC} + B_{SS}$ events at zero.

### 3.2.2 Settings

The TMVA Fisher discriminant method is performed in two steps.

- **First step:** "Training"

At this stage, the goal is to train the multivariable tool TMVA with well defined samples of signal and background. This will allow to compute the Fisher variable (section 3.2.1). Therefore, the inputs we choose are the two samples of MC-generated signal and SS-background (section 1.1).

The training is performed using the pre-implemented file `TMVAClassificationCategory.cxx`. This file is filled specifying the method we want to use (Fisher) and the variables needed to execute this method ($M$, $Q$, $p_{T_{j/\psi}}$, $p_{T_{\mu_1}}$, $p_{T_{\mu_2}}$, $pidk_1$ and $pidk_2$).

It is known that the the likelihood differences for the two pions ($pidk$) distributions are not well reproduced by MC-simulation. For this reason, $pidk_1$ and $pidk_2$ are not good input variables for the Fisher discriminants computing. We prefer to constrain them in separate cuts $pidk_1 \leq 0$ and $pidk_2 \leq 0$. The mass with the already stated cuts : $3862 \leq M \leq 3882$.

---

4We decided to chose only $p_{T_{\mu_1}}$ and $p_{T_{\mu_1}}$ and not to consider $\min(p_{T_{\mu_1}}, p_{T_{\mu_2}})$ and $\min(p_{T_{\mu_1}}, p_{T_{\mu_2}})$ or $\max(p_{T_{\mu_1}}, p_{T_{\mu_2}})$ and $\max(p_{T_{\mu_1}}, p_{T_{\mu_2}})$. Indeed, the distributions of $p_{T_{\mu_1}}$ (resp. $p_{T_{\mu_1}}$) is similar to the distribution of the minimum or the maximum of the $p_T$ of the two muons (resp. the two pions) especially because $p_{T_{\mu_1}}$ (resp. $p_{T_{\mu_1}}$) and $p_{T_{\mu_2}}$ (resp. $p_{T_{\mu_2}}$) are very correlated. In the end, it doesn’t really matter to use only $p_{T_{\mu_1}}$ and $p_{T_{\mu_1}}$ instead of the minima and maxima. Although, it saves a considerable amount of time to use simple variable instead of altered ones, particularly because it allows to use directly the data of the $n$-Tuples without having to apply any mathematical operation on them.
The variables used for the computation of the Fisher variable are: $Q$, $p_{T_{J/\psi}}$, $p_{T_{x_1}}$, and $p_{T_{\mu_1}}$. Figure 5 shows the correlations for signal and background between each variable.

![Correlation Matrix](image)

(a) Correlation Matrix for Background       (b) Correlation Matrix for Signal

Figure 5: Correlations matrices for each variable $Q$, $p_{T_{J/\psi}}$, $p_{T_{x_1}}$, $p_{T_{\mu_1}}$, $p_{idk_1}$, $p_{idk_2}$ and $M$.

As expected, the momentum of the first muon is strongly correlated to the $J/\psi$ momentum as a daughter particle for the two samples. For both signal and background, the mass has no correlation with other variables, which secures that there is no undesirable hidden mass dependence in other variable. Indeed, the variable we are performing cuts on have to be strictly independent from the mass. Otherwise, to perform cuts on mass-dependant variables amounts to cut some values on the spectrum of mass. This may badly influence our determination of the $X(3872)$ mass.

To improve the efficiency of the data processing, three preliminary transformations on the data are implemented in TMVA [12].

- **Variable normalization**: Maximum and minimum values for all input variables are determined from the training events and used to linearly scale the input variables to lie within [-1,1].

- **Decorrelation**: Linear correlations, measured in training sample, can be taken into account through computing the square-root of the covariance matrix. The linear decorrelation of the input variables is then obtained by multiplying the initial variable by the inverse of the square-root matrix. The transformations are performed separately for signal and background.
• **Transformation of the variables into Gaussian distributions:** The decorrelation method requires Gaussian distributed input variables. One may transform the variables such that their distribution becomes Gaussian.

The coefficients of the Fisher discriminant are then computed in such a way that the separation between signal and background will be maximized and the events belonging to signal or background respectively close-together confined. These coefficients (eq. 3) are depending on the way the input variables are dressed-up at the beginning. Let \( Q', p'_{T_{J/\psi}}, p'_{T_{\tau_1}} \) and \( p'_{\nu_1} \) be the variables after they have been transformed. The Fisher variable \( \mathcal{F} \) will be then:

\[
\mathcal{F} = F_0 + F_1 \cdot Q' + F_2 \cdot p'_{T_{J/\psi}} + F_3 \cdot p'_{T_{\tau_1}} + F_4 \cdot p'_{\nu_1}
\]

with:

- \( F_0 = + 8.95 \cdot 10^{-4} \)
- \( F_1 = - 1.06 \cdot 10^{-1} \)
- \( F_2 = - 2.16 \cdot 10^{-3} \)
- \( F_3 = - 4.38 \cdot 10^{-6} \)
- \( F_4 = - 7.29 \cdot 10^{-6} \)

On figure 6, the background rejection i.e. the proportion of background which is not passing through the Fisher’s cut, is represented as a function of the efficiency of the signal, i.e., the proportion of signal events passing through the Fisher’s cut. The best quality of the signal is reached when the background rejection is strong enough to get rid of a great part of the noise, but not too strong either so as to still keep some signal.

On figure 7 are plotted the significance, the purity and the efficiencies of each sample for different values of the Fisher’s cut. For a given Fisher’s cut value, the signal and background efficiencies are derived by counting the training events that pass the cut and dividing the number found by the original sample sizes.

For 585 events of signal and 5013 events [4] of background (see section 1.2) we have:

- the maximum of significance, \( \mathcal{F}_{\text{max}} = 14.9 \)
- the optimum Fisher’s cut, \( \mathcal{F} = 0.1728 \)
Figure 6: Background rejection versus signal efficiency. The greater the number of events of signal passing the cut will be, the more we will have background events hiding our signal.

Figure 7: Efficiency of the signal (blue plain line), efficiency of the background (red plain line), significance (green plain line), purity (blue dashed line), (purity-efficiency of the signal) (blue dotted curve)
• Second step : Application

Once the training is complete, one can apply the computed weights to the transformed variables of a real sample of data containing both signal and background. This is done with the pre-implemented file `TMVAClassificationApplication.cxx` using the file `TMVAClassificationCategory_Fisher.weights.xml` for the Fisher variable.

To control the smooth functioning of the tool, one can apply this process to our initial sample of SS-background to control either the result corresponds to the expected distribution of background we obtained after training or not. The distributions of background and signal (versus the Fisher’s variable) obtained after training are compared with the distribution of background whose variables have been transformed with the training weights. The distributions before and after applying the Fisher method are the same, as attested by Figure 8.

![Figure 8: Outputs of TMVA Fisher’s method for MC and SS-distributions. The filled histograms are corresponding to the results of the training versus the Fisher’s variable $F$: in blue, the SS-background, in red, the MC-signal. The black line corresponds to the result of the application of Fisher’s coefficients on the sample of SS-background. The distributions corresponding to background after training and after application of Fisher coefficients are superimposing quite well. On this graph, the events of signal and background are presented with the same normalisation. Further, one can see on figure 8 (dashed line), that imposing the cut on the Fisher’s value at 0.1728 allows to reject lots of background events and keep a significant part of the signal.](image_url)

$\int_{F_{\text{min}}}^{F_{\text{max}}} N \cdot dF = 1$ where $N = S, B$

In this case, the number of background events is equal to the number of signal events, which is not right in reality. Indeed, the distribution of signal appears very small comparing to background without normalization.

---

5These two distributions are normalized this way: $\int_{F_{\text{min}}}^{F_{\text{max}}} N \cdot dF = 1$ where $N = S, B$
3.2.3 Application

Finally, we can apply the Fisher’s cut $\mathcal{F} > 0.1728$ to our sample of real data in order to select the best amount of signal as compared with background.

(a) Invariant mass distribution of $J/\psi\pi^+\pi^-$ without Fisher’s discriminant optimization with the selective cuts of Table 3. The fit to the mass distribution is represented with a red line.

(b) Invariant mass distribution of $J/\psi\pi^+\pi^-$ with the Fisher’s discriminant optimization: $\mathcal{F} > 0.1728$ and $\text{pid}_{1,2} < 0$. The fit to the mass distribution is represented with a blue line.

Figure 9: Comparison of invariant mass distribution without (Fig. 9(a)) and with Fisher’s method for optimization (Fig. 9(b)). On the first plot, the $X(3872)$ resonance is not visible. Only the $\psi(2S)$ peak can be seen.

The final mass distribution is reported on figure 9(b) and can be compared with the
original mass distribution without cuts on the Figure 9(a).

A fit is performed on the distribution (Table 6) which gives an approximation of the energy associate to the maximum of the peak, of the width $\sigma$ of the resonance, of the number $N_{\text{events}}$ of counted events in this resonance and of the purity $P$. The results of the fit with this Fisher's variable for $\psi(2S)$ and the background are given in appendix.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Fit value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass</td>
<td>$3869.27 \pm 1.02$ MeV</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>$6.43 \pm 1.55$ MeV</td>
</tr>
<tr>
<td>$N_{\text{events}}$</td>
<td>$1248 \pm 212$</td>
</tr>
<tr>
<td>$P = S/B$</td>
<td>$0.12$</td>
</tr>
<tr>
<td>$\mathcal{I}$</td>
<td>11.6</td>
</tr>
</tbody>
</table>

Table 6: Fit results for $X(3872)$

Using Fisher’s discriminants method finally allows to optimize the isolation of enough signal in order to bring out the resonance peak corresponding to $X(3872)$. The goal has been reached: we managed to improve the significance by an appropriate selection of events. The maximum value of significance from the fit on the data is $\mathcal{I} \approx 11.6$. Indeed, it is a better result than what we got for the simple manual cuts. Although Fisher’s discriminant method seems more efficient, the measured $\mathcal{I}$ is rather lower than the predicted one (see Fig. 7). This may be due to the following issues:

- The signal is not well described by the Monte-Carlo simulation: the input values to the MC, the quantum number assumed to be $1^{++}$, the dipion mass spectrum computed assuming that the $\rho^0$ resonance dominates the decay or the intrinsic width seted to 0.314 MeV/$c^2$ may not be correct.

- The background is not only composed of combinatorial events. Misreconstructed $J/\psi \rightarrow \mu\mu$ might contribute to the total effective background as well as other sources. In this case, it may be not correct to consider the background behaving as a SS-background. De facto, the normalization (see section 1.2) we performed on the background sample will not be correct.

With these conditions and bearing in mind that the resonance peak’s resolution is quite small (see Fig 9(b)) due to the lot of background, a precise evaluation of the $X(3872)$ mass cannot be performed in such a way. Though, the Fishier discriminant method is rather a good way to select a lot of signal events.
3 OPTIMIZED SELECTION

3.3 Rectangular cuts with TMVA

3.3.1 Operational principles

The application of an ensemble of rectangular cuts on discriminating variables using TMVA is an easy way to select signal events from a mixed sample of signal and background events.

The optimisation of cuts maximises the background rejection at a given signal efficiency, and scans over the full range of the latter quantity. The cut ensemble leading to maximum significance corresponds to a particular working point on the efficiency curve, and can be easily derived after the cut optimisation scan has converged. The following file is created: TMVAClassificationCategory.Cuts.weights.xml containing the optimal set of cuts for 100 given efficiencies of background and signal [13].

3.3.2 Results

In the hundred sets of cuts proposed, we selected the ones which were in agreement with the preselections cuts. We finally set our heart on two of these selected sets on account of their properties. Table 7 sums up the advantages and drawbacks of the two.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Set 1</th>
<th>Set 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q &lt; 133$</td>
<td></td>
<td>$Q &lt; 110$</td>
</tr>
<tr>
<td>$p_{T_{J/\psi}} &gt; 2300$</td>
<td></td>
<td>$p_{T_{J/\psi}} &gt; 2700$</td>
</tr>
<tr>
<td>$p_{T_{\mu_1}} &gt; 1120$</td>
<td></td>
<td>$p_{T_{\mu_1}} &gt; 780$</td>
</tr>
<tr>
<td>$p_{T_{\pi_1}} &gt; 970$</td>
<td></td>
<td>$p_{T_{\pi_1}} &gt; 590$</td>
</tr>
<tr>
<td>$J$</td>
<td>12.4</td>
<td>17.5</td>
</tr>
<tr>
<td>$\epsilon_S/\epsilon_B$</td>
<td>150</td>
<td>30</td>
</tr>
<tr>
<td>Advantages</td>
<td>Great efficiency. Signal efficiency is 150 times greater than background one.</td>
<td>Great significance as compared with these calculated with Fisher and manual cuts.</td>
</tr>
<tr>
<td>Drawbacks</td>
<td>The significance is not so high as compared with this obtained with Set 2. Fisher’s method gives us better results for significance.</td>
<td>The efficiency is not so high as for Set 1. Signal efficiency is only 30 times greater than background one.</td>
</tr>
</tbody>
</table>

Table 7: Comparing of two set of cuts

---

6TMVA looks for all the sets of cuts possible, scanning the whole range of values that can be taken by each variables $Q$, $p_{T_{J/\psi}}$, $p_{T_{\mu_1}}$, and $p_{T_{\pi_1}}$. The program does not forbid to choose a cut for which the number of events at a given value of the considered variable is zero. For example, for the $p_{T_{J/\psi}}$, taking a cut for a value of $p_{T_{J/\psi}}$ less than 2 GeV, means to select 0 event. This type of cuts, is obviously not desirable for the selection. That the reason why we do not consider the sets proposing cuts which are not in the range established by the preselection.
ε_S (resp. ε_B) is the efficiency of the signal (resp. the background), i. e. the number of signal (resp. background) events which passed the whole sets of cuts divided by the initial number of events of signal (resp. background).

### 3.3.3 Application

We can apply these two sets of cuts on a sample of real signal. The results are presented in figures 10 for the first set of cuts and 11 for the second one.

![Figure 10](https://example.com/figure10.png)  
**Figure 10:** Invariant mass distribution of \( J/\psi \pi^+ \pi^- \) with the first set of cuts. The fit to the mass distribution is represented with a red line.

![Figure 11](https://example.com/figure11.png)  
**Figure 11:** Invariant mass distribution of \( J/\psi \pi^+ \pi^- \) with the second set of cuts. The fit to the mass distribution is represented with a blue line.

As previously, a fit is performed on these distributions and gives us the position of the peak and its width as well as the number of events corresponding to this peak and the purity of the signal (Table 8). The fitting results for \( \psi(2S) \) and the background are reported in the appendix.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Fit value for Set 1</th>
<th>Fit value for Set 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass</td>
<td>( 3871.39 \pm 0.90 ) MeV</td>
<td>( 3871.95 \pm 0.46 ) MeV</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>( 3.29 \pm 0.73 ) MeV</td>
<td>( 3.02 \pm 0.69 ) MeV</td>
</tr>
<tr>
<td>( N_{\text{events}} )</td>
<td>( 62 \pm 13 )</td>
<td>( 329 \pm 44 )</td>
</tr>
<tr>
<td>( \mathcal{P} = S/B )</td>
<td>0.76</td>
<td>0.39</td>
</tr>
</tbody>
</table>

Table 8: Results of the fits for each set of cuts
Both sets are compatible with the actual results. The first one provides a better purity but the statistic is not sufficient enough which explains the bad resolution of the peak and the high statistical error.

The second set gives the best results in term of resolution: the peak at the $X(3872)$-resonance is thinner. For this selected sample, the significance is $S \approx 10$. It is less than what we have been able to derived from the TMVA rectangular cuts on the training samples. This difference is probably due to the same factors discussed at section 3.2.3. The fit on the data provides the same value for the statistical error as in the note [4]. The resolution of the peak is rather low.

To improve these results, it is suggested to increase the size of the samples to deal with more statistics to get a better resolution with 2011 data. It will be then possible to determine the $X(3872)$ mass with a precision better than what was previously made ([4]), more particularly thanks to the good cutting of background events.

**Conclusion**

We have been testing three different methods to optimize the significance of our samples and applied them on real data. The results have been conclusive for the optimized rectangular cuts with TMVA [12]. This method allowed us to keep a maximum of signal and cut a great amount of noise. For a well chosen set of cuts, we managed to obtain quite good significance after fitting on the data ($S = 10$) for a high purity ($P \approx 0.4$). We have been computing the following mass for $X(3872) = 3871.95 \pm 0.46$ MeV/c$^2$ which is in keeping with the actual other mass measurements for this particle$^7$.

In this case, the computation of each optimal cut has been done for each variable independently, without taking into account that the significance may be better by making the cuts values evolve simultaneously. The TMVA rectangular cuts try to counteract this effect by testing randomly a lot of combinations of possible cuts and looking for the most appropriate at a given efficiency. In the manual cuts, we were constrained to maintain all the cuts fixed when scanning over one cutting variable, despite the fact that a single cut on $Q$ allows to get read of the most part of the background. The Fisher’s discriminant method has not this disadvantage because it is based on a linear combination of each cutting variable. Although the significance results are better than those we got with manual cuts, the resolution is not good enough and cannot provide a satisfying estimation of the $X(3872)$ mass. This could be improved by trying to find the best non-correlated variables to use in input.

Several other optimization methods such as likelihood estimators or neural networks are currently used to improve the quality of the results.

$^7$Actual mean estimation of the $X(3872)$ mass from Particle Data Group [14]: 3871.56 ± 0.22 MeV.
Appendix
Optimization of the $X(3872) \rightarrow J/\psi \pi^+ \pi^-$ selection at LHCb

June 16, 2011
1 Manual cuts with ROOT

Here follows the C++ code we used to implement the determination of the best cuts to perform on each variable.

```cpp
// ----------------------------------------BEGINNING OF THE LOOP----------------------------------------
for(int n=0; n<mBin_cut; n++){
  double cut = cut_min + n*(cut_max-cut_min)/mBin_cut;
  char cutPiPi_SS[600];

  //Definition of the cuts. Variation on Q
  sprintf(cutPiPi_SS, "Q < %i && PT_JPSI > 3500 && TMath::Min(pt_pi1, pt_pi2) > 500 && TMath::Min(pt_mu1, pt_mu2) > 1000 && pTkl < 0 && pTk2 < 0 && momentumM < 3676 && momentumM < 3696", cut);

  TCut TCut_cut;
  TCut_cut = cutPiPi_SS;
  nCan_data_SS->Reset();
  nCan_data_psi->Reset();
  DecayTreePiPi_SS->Draw(\"nCan\">>nCan_data_SS", TCut_cut);
  DecayTreePiPi_psi->Draw(\"nCan\">>nCan_data_psi", TCut_cut);

  //Counting of the number of events for each sample and normalization
  double S = nCan_data_psi->GetEntries();
  S=S*0.04;
  double B = nCan_data_SS->GetEntries();
  B=B*0.24;
  double rap = S/TMath::Sqrt(S+B);

  //Filling of the histogram for the significance
  SoverB->Fill(cut, rap);
  nCan_data_SS->Reset();
  nCan_data_psi->Reset();
}
//----------------------------------------END OF THE LOOP----------------------------------------

c1 = new TCanvas("c1", ", 500,500);
c1->cd();
SoverB->Draw();

double maxi = SoverB->GetMaximum();
double bin_cut_rough = (300-cut_min)*mBin_cut/(cut_max-cut_min);
double rough = SoverB->GetBinContent(bin_cut_rough);

for(int i=1; i <mBin_cut; i++){
  double bina = SoverB->GetBin(i);
  double binContent = SoverB->GetBinContent(bina);

  if(binContent == maxi) {
    cout << "Optimum cut : " "cut_min + bina*(cut_max-cut_min)/mBin_cut << endl;
    cout << "S_max = " "maxi << endl;
  }
}
//----------------------------------------IMPROVING WITH REGARD TO THE ROUGH CUTS----------------------------------------
cout << "Improving in percent : " "100*(maxi-rough)/(maxi+rough)<<%" << endl;
```
2 RESULTS OF THE FIT FOR $\psi(2S)$ AND BACKGROUND

2 Results of the fit for $\psi(2S)$ and background

2.1 Manual cut on Q

<table>
<thead>
<tr>
<th>Variable</th>
<th>Fit value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass</td>
<td>$3686.15 \pm 0.06$ MeV</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>$2.59 \pm 0.06$ MeV</td>
</tr>
<tr>
<td>$N_{\text{events}}$</td>
<td>$3703 \pm 75$</td>
</tr>
<tr>
<td>$\mathcal{P} = S/B$</td>
<td>$2.34$</td>
</tr>
</tbody>
</table>

Table 1: Fit results for $\psi(2S)$ with a manual cut on Q

<table>
<thead>
<tr>
<th>Variable</th>
<th>Fit value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B$</td>
<td>$42413 \pm 218$ MeV</td>
</tr>
<tr>
<td>$c_0$</td>
<td>$5.17 \pm 0.16$ MeV</td>
</tr>
<tr>
<td>$c_1$</td>
<td>$0.10 \pm 0.0$</td>
</tr>
<tr>
<td>$c_2$</td>
<td>$-1.2 \cdot 10^{-5} \pm 0$</td>
</tr>
</tbody>
</table>

Table 2: Fit results for background with a manual cut on Q

2.2 Fisher discriminant method

<table>
<thead>
<tr>
<th>Variable</th>
<th>Fit value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass</td>
<td>$3686.08 \pm 0.05$ MeV</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>$2.59 \pm 0.05$ MeV</td>
</tr>
<tr>
<td>$N_{\text{events}}$</td>
<td>$6162 \pm 108$</td>
</tr>
<tr>
<td>$\mathcal{P} = S/B$</td>
<td>$1.26$</td>
</tr>
</tbody>
</table>

Table 3: Fit results for $\psi(2S)$ with Fisher’s method
2 RESULTS OF THE FIT FOR ψ(2S) AND BACKGROUND

<table>
<thead>
<tr>
<th>Variable</th>
<th>Fit value</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>168919 ± 467 MeV</td>
</tr>
<tr>
<td>$c_0$</td>
<td>6.08 ± 0.09 MeV</td>
</tr>
<tr>
<td>$c_1$</td>
<td>0.10 ± 0.00</td>
</tr>
<tr>
<td>$c_2$</td>
<td>$-1.2 \cdot 10^{-5} ± 0$</td>
</tr>
</tbody>
</table>

Table 4: Fit results for background with Fisher’s method

2.3 TMVA rectangular cuts

<table>
<thead>
<tr>
<th>Variable</th>
<th>Fit value for Set 1</th>
<th>Fit value for Set 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass</td>
<td>3686.3 ± 0.2 MeV</td>
<td>3686.2 ± 0.1 MeV</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>3.01 ± 0.17 MeV</td>
<td>2.67 ± 0.07 MeV</td>
</tr>
<tr>
<td>$N_{events}$</td>
<td>350 ± 20</td>
<td>1890 ± 50</td>
</tr>
<tr>
<td>$P = S/B$</td>
<td>6.41</td>
<td>4.13</td>
</tr>
</tbody>
</table>

Table 5: Fit results for ψ(2S) with rectangular cuts

<table>
<thead>
<tr>
<th>Variable</th>
<th>Fit value for Set 1</th>
<th>Fit value for Set 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>1311 ± 39 MeV</td>
<td>14198 ± 128 MeV</td>
</tr>
<tr>
<td>$c_0$</td>
<td>3.98 ± 0.92 MeV</td>
<td>4.53 ± 0.29 MeV</td>
</tr>
<tr>
<td>$c_1$</td>
<td>0.19 ± 0.00</td>
<td>0.10 ± 0.00</td>
</tr>
<tr>
<td>$c_2$</td>
<td>$-1.2 \cdot 10^{-5} ± 0$</td>
<td>$-1.2 \cdot 10^{-5} ± 0$</td>
</tr>
</tbody>
</table>

Table 6: Fit results for background with rectangular cuts
References


